NONPARAMETRIC ESTIMATION OF THE CUMULATIVE INTENSITY FROM DOUBLY CENSORED DATA. ASYMPTOTIC THEORY.

by

SVEN OVE SAMUELSEN

ABSTRACT: Nonparametric estimators of the cumulative intensity and the survival function, making use of all available information when the data is doubly-censored, are introduced. The estimators are analogous to the Kaplan-Meier P-L-estimator. The estimator of the cumulative intensity is shown to be consistent and by use of the martingale-central-limit-theorem its normalization is shown to converge to a gaussian process. The same results are valid for the estimator of the survival function. A simulation study comparing the estimators with other natural choices of estimators is presented.

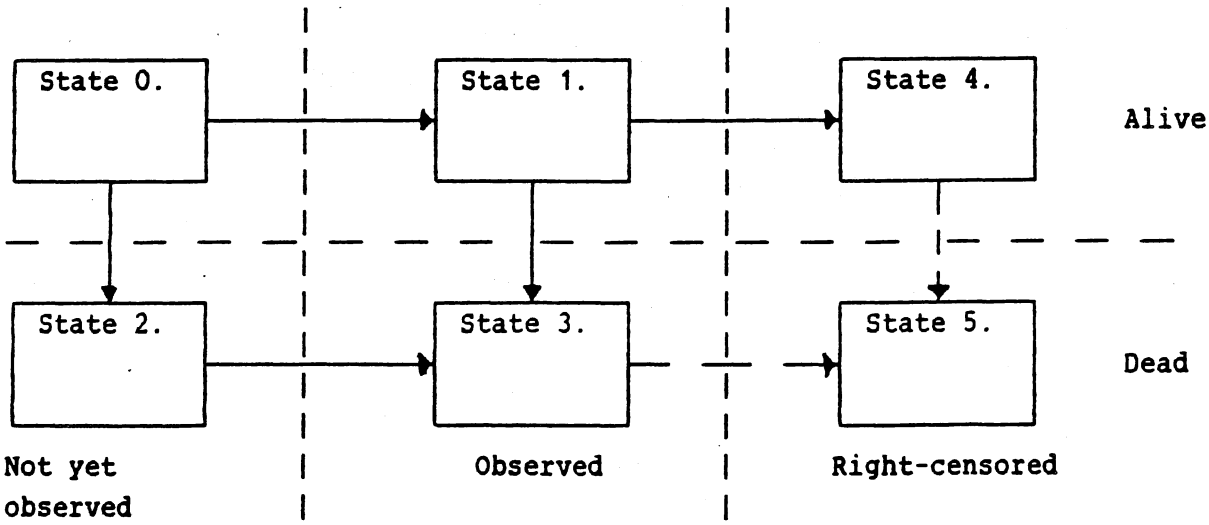
KEY-WORDS: Doubly-censored data, nonparametric estimation, cumulative intensity, Nelson-Aalen plott, Kaplan-Meier P-L-estimator, counting process/martingale theory, asymptotic theory, simulation study.

1. Introduction.

In this paper we want to study nonparametric estimatiors of the cumulative intensity of a life-time T when the data is doubly-censored. By doubly-censored we mean that to each life-time T_i $i=1,\ldots,n$ there corresponds a "window" of observation $\langle L_i, U_i]$ so that T_i is accuratly observed if $L_i < T_i \leq U_i$, left-censored if $T_i \leq L_i$ and right censored if $T_i > U_i$. Observation will take place on a fixed time-period, say [0,1]. We will need that L_i is always observed.

The complete data corresponding to individual i (i.e. to the data (L_i, T_i, U_i)) may be described as a process on the compartments 0.-5. in figur 1.1 where transitions are allowed in the directions pointed out by the arrows. Here L_i=time until the individual comes under observation=time until the process first reaches one of the states 1. or 3. Similarly U_i=time until (eventual) right-censoring= time until the process first reaches one of the states 4. or 5. and T_i=lifetime of individual i=time until the first transition to one of the states 2., 3. or 5.

Figur 1.1



From the observed data we are able to tell if an individual is in state 0. or state 1., but not which of these states she is in. Further we are always able to tell if the individual is in state 1. and if a transition from 1. to 3., from 1. to 4. or from 2. to 3. has occured before a given time t.

The individual processes will generally be semi-markovian (and markovian if the intensity for right-censoring is a function only of time and does not depend on the time spent in state 1. or 3.).

Papers treating related models with nonparametric methods are Peto[1973] and Turnbull[1974]. These both give examples on how such situations may occur. From our point of view the paper of Turnbull is of particular interest. He gives an iterative estimation procedure which he shows yield MLE for the distribution function (See Turnbull[1974] and [1976]). This estimator will later be referred to as the Turnbull-estimator. It is an extension of the self-consistency estimator of Efron[1967] and may be shown to be an instance of the E-M-principle of Dempster et. al. [1977]. The estimator may be generalized to a wide class of incomplete data problems (see Turnbull[1976]).

Our intention for this paper was to extend the results of Turnbull to the case of continuos lifetime distributions. However this has its complications since 1) the Turnbull estimator is given as a limit of an iterative algorithm and not as an explicit formula and 2) the covariance matrix (i.e. the inverse of the information matrix) of the estimator has not, as far as we know, a nice form.

Hence we have found it convenient to impose the restriction (unnecessary in Turnbull[1974]) that L_i is observed (whenever $L_i<1$). This makes it possible to construct a P-L-estimator (see Kaplan&Meier[1958]) with nice and wellknown properties from parts of the datamaterial . In turn using this estimator as a startpoint for the Turnbull-estimator and iterating just once we get an estimator using all relevant information in a sound way, thus containing the "spirit" of the Turnbull-estimator.

This one-step-estimator is, like the P-L-estimator, a product. We may, analogues to the Nelson-Aalen-estimator, construct a corresponding estimator for the cumulative intensity on the form of a sum. We will focus upon this estimator; partly because sums are somewhat more easy to handle than products, partly because intensities are natural parameters in survival analysis. Corresponding results for the estimator analogous to the P-L-estimator can be found using the same technique as Breslow&Crowley[1974] and Aalen[1978a] (i.e by Taylor expansions.)

We will now introduce some notation and asumptions that will be used throughout the paper.

Formally our data are originated from the complete data set $\{(L_i,T_i,U_i);1,\dots,n\}$ and we observe $\{(L_i,X_i,\partial_i);i=1,\dots,n\}$ where

$$X_i=\max(\min(T_i,U_i),L_i) \tag{1.1}$$

and

$$\partial_i=\Delta\{X_i=U_i<T_i\}. \tag{1.2}$$

$\Delta\{A\}$ equals 1 or 0 according to whether the expression A is true or not. In case the truth of A varies with time Δ is considered as a function.

Notice that $X_i=U_i$ if T_i is right-censored, $X_i=T_i$ if T_i is accuratly observed and $X_i=L_i$ if T_i is left-censored. δ_i is indicating right-censoring.

We will assume

1) The (L_i,T_i,U_i) are independent and identically distributed.
2) (L_i,U_i) is independent of T_i for each i.
3) $L_i< U_i$ with probability 1.
4) $P(L_i=0) = \lambda > 0$ and $P(L_i<0)=0$.
5) $\inf_{t\epsilon[0,1]} P\{ (L_i<t)\cap [(T_i>t) \cup (U_i>t)] \} > 0$.
6) The T_i, $L_i|L_i>0$ and U_i are absolutely continous variables on $[0,\infty\rangle$.
7) $P(L_i>1) = 0$.

In Table 1.1 we give notation for distribution functions of T_i, $L_i|L_i>0$ and U_i.

Table 1.1

Variable	c.d.f.	density	surv.func.	intensity	cum.intensity
T_i	$F(t)$	$f(t)$	$S(t)$	$\alpha(t)$	$\beta(t)$
$L_i\|L_i>0$	$G(t)$	$g(t)$	$S_G(t)$	$\gamma(t)$	-
U_i	$H(t)$	$h(t)$	$S_H(t)$	-	-

We conclude this section by an outline of the rest of the paper. In the next section we introduce counting processes for the (complete and censored) data and then estimators of the cumulative intensity and survival function are introduced. We end the section by relating martingales to the counting processes. The asymptotic result for the estimator of the cumulative intensity is given in section 3 and 4, consistency in section 3 and weak convergence in section 4. The corresponding results for the estimator of the survival function are reviewed in section 5. In section 6 we present a simulation study of doubly-censored data and make comparisons between different possible estimators. The paper is concluded by a discussion.

2. A counting process formulation and the "one-step-estimator".

For each of the n individuals we define

$Y_{ij}(t)=\Delta\{$ individual i is in state j at t-$\}$

$N_{ijk}(t)=\Delta\{$transition of individual i from state j to state k in $[0,t]\}$

where $j,k=0,\dots,5$.

On an accumulated level we have the corresponding counting processes and risk-sets:

$$Y_j(t)=\sum_{i=1}^{n} Y_{ij}(t)$$

and

$$N_{jk}(t)=\sum_{i=1}^{n} N_{ijk}(t)$$

for the same j and k.

Not all $Y_{ij}(t)$ and $N_{ijk}(t)$ are observed. For instance $Y_{i0}(t)$ and $Y_{i2}(t)$ are not known apart from their sum. Likewise a jump of $N_{i02}(t)$ will not be observed when it occurs.

We will also need the following (partially) unknown quantities

$$Y(t)=Y_0(t)+Y_1(t) \tag{2.1}$$

and

$$N(t)=N_{02}(t)+N_{13}(t) \tag{2.2}$$

which are the risk set for and counting process of the deaths that are not right censored.

Suppose N(t) and Y(t) were known, then the P-L-estimator of S(t) would be

$$\tilde{S}(t)=\prod_{s\leq t}[1-Y(s)^{-1}dN(s)] \tag{2.3}$$

Similarily the Nelson-Aalen estimator of $\beta(t)$ would be given by

$$\tilde{\beta}(t) = \int_0^t Y(s)^{-1}dN(s) \tag{2.4}$$

The representation of these estimators as respectively product-integrals and (sum-)integrals is explained in e.g. Andersen&Borgan[1985].

Allthough Y(t) and N(t) are unknown they may, as we shall see, be estimated. It is reasonable to estimate S(t) and $\beta(t)$ by simply replacing Y(t) and N(t) in (2.3) and (2.4) by such estimates.

Let $\mathcal{F}= \sigma\,\{(L_i,X_i,\delta_i);\ i=1,\dots\dots,n\}$ i.e. the σ-algebra generated by the observations. Then we have

$$E[Y(t)|\mathfrak{F}]=Y_1(t)+[N_{01}(1)-N_{01}(t-)]+\int_t^1 [S(t)-S(s)]\ [1 - S(s)]^{-1}dN_{23}(s) \quad (2.5)$$

and

$$E[N(t)|\mathfrak{F}]= N_{13}(t) + N_{23}(t) + [1-S(t)] \int_t^1 [1-S(s)]^{-1}dN_{23}(s) \quad (2.6)$$

Expression (2.5) may be verified the following way. The term $Y_1(t)$ enters since $Y(t)=Y_1(t)+Y_0(t)$ and $Y_1(t)$ is known (measurable) in $\mathfrak{F}$. $N_{01}(1)-N_{01}(t-)$ is the sum of individuals that enters the observed group after time t, and hence these individuals neccesarily are among the $Y_0(t)$ that are not yet observed but alive at t. The probability that an individual whose lifetime is left-censored at $s > t$ is alive at t $[S(t)-S(s)]\ [1-S(s)]^{-1}$ and then integrating over the leftcensorings after t we obtain the third term in (2.5). The verification of (2.6) is similiar to the preceeding.

In (2.5) and (2.6) only S(t) is unknown. However as a direct consequence of the assumption that L_i is observed (whenever $L_i<1$) we may estimate S(t) by the P-L-estimator based on the subsample under direct observation, i.e.

$$\hat{S}(t) = \prod_{s<t} [\ 1 - Y_1(s)^{-1}dN_{13}(s)\ ] \quad (2.7)$$

Similarily define a Nelson-Aalen estimator of the cumulative intensity

$$\hat{\beta}(t) = \int_0^t Y_1(s)^{-1}dN_{13}(s) \quad (2.8)$$

Under our assumptions $\hat{S}(t)$ is asymptotically unbiased, strongly consistent and $\sqrt{n}\ (\ \hat{S}(t) - S(t)\ )$ tends to a gaussian process as $n \to \infty$. These facts are proved under slightly different conditons in Gill[1980] and Aalen&Johansen[1978] and may also easily be proved by results presented later in the paper. The main reason they hold is assumption 4) which is equivalent to $\inf_{t\in[0,1]} E\ Y_1(t)/n\ > 0$.

If there are no left-censored deaths prior to the first observed death we may now estimate Y(t) and N(t) by simply inserting $\hat{S}(t)$ into the left-hand side of (2.5) and (2.6). If however $\hat{S}(t) = 1$ for a t with $dN_{23}(t) > 0$ we will simply ignore such a jump. Ideally such information should have been used in one way or another, but as our goal is primary asymptotics this will be of little concern. Due to this omission it will also be convenient to ignore jumps of $N_{01}(t)$ when $\hat{S}(t) = 1$.

Set

$$J_{jk}(t)= \Delta \{ N_{jk}(t)>0 \}$$

Then by interpreting 0/0=0, our estimators of Y(t) and N(t) become

$$Y^*(t)=Y_1(t)+\int_t^1 J_{13}(s)dN_{01}(s)]+\int_t^1 [\hat{S}(t-)-\hat{S}(s)] [1 - \hat{S}(s)]^{-1} J_{13}(s)dN_{23}(s) \quad (2.9)$$

and

$$N^*(t)= N_{13}(t) + J_{13}(t)N_{23}(t) + [1-\hat{S}(t)] \int_t^1 [1-\hat{S}(s)]^{-1} J_{13}(s)\, dN_{23}(s) \quad (2.10)$$

According to a preceeding remark we should now estimate S(t) and $\beta(t)$ by respectively

$$S^*(t)=\prod_{s\leq t}(1-Y^*(s)^{-1}dN^*(s)) \quad (2.11)$$

and

$$\beta^*(t)= \int_0^t Y^*(s)^{-1}dN^*(s) \quad (2.12)$$

In the phrasing of the E-M-algorithm formulas (2.9) and (2.10) corresponds to the E-step, while formula (2.11) gives the M-step. Thus by replacing S(t) by $S^*(t)$ in (2.9) and (2.10) and inserting these new estimates of Y(t) and N(t) in (2.11) the second iteration for the estimator of the life-time-distribution arises. By repeating this process until convergence we get the Turnbull-estimator (when F(t) has jumps only at the jumps of $N_{13}(t)$.)

By defining

$$Y^{**}(t)= Y_1(t) + \hat{S}(t-) \int_t^1 [1-\hat{S}(s)]^{-1} J_{13}(s)\, dN_{23}(s)]$$

we get

$$dN^*(t) = Y^{**}(t)\, Y_1(t)^{-1}\, dN_{13}(t)$$

since $-d\hat{S}(t) = \hat{S}(t-)\, Y_1(t)^{-1}dN_{13}(t)$

Hence (2.12) becomes

$$\beta^*(t) = \int_0^t Y^{**}(s)\, Y^*(s)^{-1}\, Y_1(s)^{-1}\, dN_{13}(s)$$

We shall later see that $Y^{**}(t)$ is also a reasonable estimator for Y(t), hence motivating the estimator $\beta^*(t)$.

We will now see how martingales can be related to the counting processes. The martingales will mainly be used in proving weak convergence, but a few steps in the in the determination of the constistency may also be simplified by martingale results. For an explanation of the concepts in the multivariate counting process/martingale framework see Aalen[1978b] or Gill[1980].

Let

$\mathcal{F}_t = \sigma \{ n, Y_1(0), N_{01}(s), N_{13}(s), N_{14}(s), N_{23}(s); 0 \leq s \leq t \}$

i.e. the σ-algebra generated by the observations up to time t.

and

$\mathcal{G}_t = \sigma \{ n, Y_1(0), N_{01}(s), N_{13}(s), N_{14}(s), N_{23}(s), N_{02}(s); 0 \leq s \leq t \}$

Hence $\mathcal{G}_t$ also records the transitions from state 0. to state 2. and clearly $\mathcal{F}_t \subset \mathcal{G}_t$.

We will also define

$$\Lambda'_{13}(t) = \Lambda_{13}(t) = \int_0^t Y_1(s)\,\alpha(s)\,ds$$

$$\Lambda'_{23}(t) = \int_0^t Y_2(s)\,\gamma(s)\,ds$$

$$\Lambda_{23}(t) = \int_0^t (Y_0(s)+Y_2(s))\,(1-S(s))\,\gamma(s)\,ds$$

$$\Lambda'_{01}(t) = \int_0^t Y_0(s)\,\gamma(s)\,ds$$

$$\Lambda_{01}(t) = \int_0^t (Y_0(s)+Y_2(s))\,S(s)\,\gamma(s)\,ds$$

Note that while $Y_0(s)$ and $Y_2(s)$ are measurable in $\mathcal{G}_{s-}$ only the sum $Y_0(s)+Y_2(s)$ is measurable in $\mathcal{F}_{s-}$. Since all $\Lambda'_{jk}(t)$ and $\Lambda_{jk}(t)$ are (left-)continous we have the $\Lambda_{jk}(t)$ predictable in $\mathcal{F}_{t-}$ and the $\Lambda'_{jk}(t)$ predictable in $\mathcal{G}_{t-}$.

Finally set

$$M_{jk}(t) = N_{jk}(t) - \Lambda_{jk}(t)$$

and

$$M'_{jk}(t) = N'_{jk}(t) - \Lambda'_{jk}(t)$$

for jk=13,01,23. Then we have

<u>Lemma 2.1</u>

$M_{13}(t)$,$M_{01}(t)$ and $M_{23}(t)$ are orthogonal and quadratic integrable martingales with respect to the family of σ-algebras $\{ \mathcal{F}_t ; 0 \leq t \leq 1 \}$.

<u>Proof of lemma 2.1</u>

That $M_{jk}(t)$ is a martingales with respect to $\{ \mathcal{F}_t \}$ may also be stated that $\Lambda_{jk}(t)$ is the compensators of $N_{jk}(t)$ with respect to $\{ \mathcal{F}_t \}$. From Gill[1980] we have that $\Lambda'_{13}(t)$, $\Lambda'_{01}(t)$ and $\Lambda'_{23}(t)$ are the compensators of respectively $N_{13}(t)$, $N_{01}(t)$ and $N_{23}(t)$ with respect to ($\mathcal{G}_t$). Using the innovation theorem (Th.3.4 in Aalen[1978b]) we then get

$$\Lambda_{jk}(t) = E [\Lambda'_{jk}(t) \mid \mathcal{F}_t]$$

But for $t \geq s$ we have

$$E [Y_1(s) \mid \mathcal{F}_t] = Y_1(s)$$

$$E [Y_0(s) \mid \mathcal{F}_t] = (Y_0(s) + Y_2(s)) S(s)$$

$$E [Y_2(s) \mid \mathcal{F}_t] = (Y_0(s) + Y_2(s)) F(s)$$

thus establishing the wanted compensators with respect to $\mathcal{F}_t$.
That the $M_{jk}(t)$'s are orthogonal follows from the fact that the $\Lambda'_{jk}(t)$'s are continous and that they are quadratic integrabel just mean that $E [M_{jk}(t)]^2 = E \Lambda_{jk}(t) < \infty$.=

3. Consistency.

In this section we will prove the following theorem:

Theorem 3.1

Under assumptions 1)-6) in section 1 we have

a) for all $\varepsilon > 0$

$$\sup_{t \in [\varepsilon,1]} |(\beta^*(t) - \beta^*(\varepsilon)) - (\beta(t) - \beta(\varepsilon))| \xrightarrow{a.s} 0$$

b) $$\sup_{t \in [0,1]} | \beta^*(t) - \beta(t) | \xrightarrow{P} 0$$

The theorem is split in two parts because difficulties occur in proving the convergence of $\beta^*(t)$ when t is close to 0. These difficulties are belived to be technical, but are solved only in the case of convergence in probability. The results are proved by the help of Lemma A.1, Lemma A.2 and Lemma A.3 in Appendix A.

Proof of theorem 3.1

Let $y_1(t) = E[\, n^{-1} Y_1(t) \,]$. Then by the Glivenko-Cantelli theorem we have

$$n^{-1} N_{13}(t) \xrightarrow{a.s} E[\, n^{-1}_{13} N \ (t) \,] = \int_0^t y_1(s)\, \alpha(s)\, ds$$

and

$$n^{-1} Y_1(t) \xrightarrow{a.s} y_1(t)$$

uniformly for $t \in [0,1]$.

By lemma A.1 we now get $\int_0^t Y_1(s)^{-1} dN_{13}(s) \xrightarrow{a.s} \beta(t)$ uniformly on [0,1]. This implies, by lemma 3.2.1.(iv) in Gill[1980], that $\hat{S}(t) \xrightarrow{a.s} S(t)$ uniformly on [0,1].

We may (for all $\varepsilon > 0$) write (see (2.21))

$$\beta^*(t) - \beta^*(\varepsilon) = \int_\varepsilon^t \varphi^*(s)\, Y_1(s)^{-1} dN_{13}(s).$$

where

$$\varphi^*(s) = Y^{**}(s) \,/\, Y^*(s)$$

and hence by lemma A.1, part a) of the theorem will follow if we show $\varphi^*(t) \xrightarrow{a.s} 1$ uniformly on $[\varepsilon,1]$ for all $\varepsilon > 0$ and correspondingly b) follows from $\varphi^*(t) \xrightarrow{P} 1$ uniformly on [0,1].

Furthermore (by Glivenko-Cantelli)

$$n^{-1} N_{01}(t) \xrightarrow{a.s} (1-\lambda) \int_0^t S(s)\, g(s)ds$$

and

$$n^{-1} N_{23}(t) \xrightarrow{a.s} (1-\lambda) \int_0^t F(s)\, g(s)ds$$

uniformly on [0,1].

Using lemma A.1 once more we now get

$$n^{-1} \int_t^1 [\, 1 - \hat{S}(s)\,]^{-1} J_{13}(s)\, dN_{23}(s) \xrightarrow{a.s} (1-\lambda) \int_t^1 g(s)\, ds \qquad (3.1$$

and

$$n^{-1} \int_t^1 \hat{S}(s)\, [\, 1 - \hat{S}(s)\,]^{-1} J_{13}(s)\, dN_{23}(s) \xrightarrow{a.s} (1-\lambda) \int_t^1 S(s)\, g(s)\, ds \qquad (3.2$$

uniformly on $[\varepsilon,1]$ for each $\varepsilon>0$.

Consequently

$$\varphi^*(t) \xrightarrow{a.s} \frac{y_1(t)+(1-\lambda)S(t)\int_t^1 g(s)ds}{y_1(t)+(1-\lambda)\int_t^1 S(s)g(s)ds+(1-\lambda)\int_t^1 (S(t)-S(s))g(s)ds} = 1$$

uniformly on $[\varepsilon,1]$ for all $\varepsilon>0$ yielding theorem 3.1.a.

Note however that since for all n there exist $\eta>0$ so that $[\, 1 - \hat{S}(\eta)\,]^{-1} = \infty$ the expressions $[1-\hat{S}(t)]^{-1}$ and $\hat{S}(t)[1-\hat{S}(t)]^{-1}$ will not converge uniformly on [0,1] and hence lemma A.1 can not be used to show (3.1) and (3.2) on all of [0,1]. However lemma A.3 states that (3.1) and (3.2) will be valid also for $\varepsilon=0$ if we replace convergence almost certain with convergence in Probability. Hence $\varphi^*(t) \xrightarrow{P} 1$ uniformly on [0,1] and by the use of lemma A.1 again part b) of Theorem 3.1 hold.

Section 4. Weak Convergence.

In this section we will pursue the limiting process of $\sqrt{n}(\beta^*(t)-\beta(t))$. As was the case with the proof of consistency of $\beta^*(t)$, there are some technical problems involved in showing the convergence for t close to zero. In the main text we will therefore only give the the proof that $\sqrt{n}\,[\,(\beta^*(t)-\beta^*(\varepsilon)) - (\beta(t)-\beta(\varepsilon))\,]$ will converge weakly to a gaussian process on $D[\varepsilon,1]$ for each $\varepsilon>0$. The main result, that this also hold for $\varepsilon=0$, is only stated and its proof may be found in the appendix. In the end of the section we discuss the estimation of the covariance of the limit-process of $\sqrt{n}(\beta^*(t)-\beta(t))$.

First some comments about general notation. In some cases functions or processes $\varphi_\varepsilon(t)$ (denoted with subscribt ε) are not explicit defined. In such case $\varphi_\varepsilon(t)=\varphi(t)-\varphi(\varepsilon)$ for functions or processes $\varphi(t)$ that have been defined. Also, if the function $\varphi(t)$ has not been defined, then it can be interpreted as $\lim_{\varepsilon\to 0} \varphi_\varepsilon(t)$ for some φ_ε that has been defined. Furthermore, if not otherwise stated, all limits are taken as $n \to \infty$. Also the terms convergence in probability of a sequence of processes $\{\phi_n\}$ on $D(A)$ to ϕ (where A is an interval) and that $\sup_{t\in A} |\,\phi_n(t) - \phi(t)\,| \xrightarrow{P} 0$ will mean exactly the same thing since all limits ϕ may be found to lie in $C(A)$ with probability 1.

Let

$$H_1^u(t) = \Delta(t\leq u) - (1-\lambda) \int_0^u \int_{s\vee t}^1 S(r)\,(1-S(r))^{-1}\, g(r)\, dr\; y^*(s)^{-1} \alpha(s) ds \tag{4.1}$$

$$H_2^u(t) = \int_0^{t\wedge u} y^*(s)^{-1}\, \alpha(s)\, ds \tag{4.2}$$

and

$$c(u,t) = \int_0^1 H_1^t(s)\, H_1^u(s)\, y_1(s)^{-1} \alpha(s)\, ds + (1-\lambda) \int_0^1 H_2^t(s)\, H_2^u(s)\, g(s)\, ds \tag{4.3}$$

Then the main result may be stated

Theorem 4.1

Let $V(u)$ be a gaussian process on $D[0,1]$ with $EV(u) = 0$ and covariance function $E\,V(t)\,V(u) = c(t,u)$ given by (4.3). Then

$$\sqrt{n}\,(\,\beta^*(u) - \beta(u)\,) \xrightarrow{D} V(u)$$

on $D[0,1]$.

Let furthermore

$$c_\varepsilon(u,t) = \int_0^1 H_{1\varepsilon}^t(s)\, H_{1\varepsilon}^u(s)\, y_1(s)^{-1}\alpha(s)ds + (1-\lambda)\int_\varepsilon^1 H_{2\varepsilon}^t(s)\, H_{2\varepsilon}^u(s)\, g(s)ds \quad (4.4)$$

Then the intermediate result we shall prove here is

Proposition 4.1

Let $V_\varepsilon(u)$ be a gaussian process on $D[\varepsilon,1]$ where $\varepsilon>0$ with $EV(u)=0$ and covariance function $EV_\varepsilon(t)V_\varepsilon(u) = c_\varepsilon(t,u)$ given by (4.4). Then on $D[\varepsilon,1]$

$$\sqrt{n}\,(\beta_\varepsilon^*(u) - \beta_\varepsilon(u)) \xrightarrow{D} V_\varepsilon(u)$$

To prove Proposition 4.1 we will need two helping results stated below in Lemma 4.1 and Lemma 4.2. The second of these lemmas give a sequence of processes $\{V_{n\varepsilon}\}$ which is asymptotically equivalent with $\sqrt{n}(\beta_\varepsilon^*(u)-\beta_\varepsilon(u))$, i.e. $V_{n\varepsilon}(u)-\sqrt{n}(\beta_\varepsilon^*(u)-\beta_\varepsilon(u)) \xrightarrow{P} 0$ on $D[\varepsilon,1]$. The first lemma give the simultaneous limiting processes of two processes, from which the limiting process of $V_{n\varepsilon}$ may be determined. This lemma is placed first since it will also be necessary when we will prove lemma 4.2.

Define

$$Z_1^n(t) = \sqrt{n}\,[\int_0^t Y_1(s)^{-1}\, dN_{13}(s) - \int_0^t \Delta(Y_1(s) > 0)\, \alpha(s)\, ds] \quad (4.5)$$

and

$$Z_{2\varepsilon}^n(t) = 1/\sqrt{n}\,\{\int_0^t \Delta(s>\varepsilon)\, J_{13}(s)\, [S(s)\,(1-S(s))^{-1}dN_{23}(s) - dN_{01}(s)]\} \quad (4.6)$$

Then we have

Lemma 4.1

If $\varepsilon>0$ we have on $D[0,1]^2$ $(Z_1^n, Z_{2\varepsilon}^n) \xrightarrow{D} (Z_1, Z_{2\varepsilon})$

where $Z_1(t)$ and $Z_{2\varepsilon}$ are two independent gaussian processes with $EZ_1(t)=0$, $EZ_{2\varepsilon}(t)=0$ for $0\leq t\leq 1$ and with covariance functions

$$EZ_1(u)\, Z_1(t) = \int_0^{u\wedge t} y_1(s)^{-1}\, \alpha(s)\, ds$$

and

$$EZ_{2\varepsilon}(u)\, Z_{2\varepsilon}(t) = (1-\lambda)\int_0^{u\wedge t} \Delta(s>\varepsilon)\, S(s)\, (1-S(s))^{-1}\, g(s)\, ds$$

Comment: For some (natural) choices of $S(s)$ and $g(s)$ (e.g. when $g(s)$ and $S(s)$ represents the same distribution i.e. $g(s)ds=-dS(s)$) will the covariancefunction of $Z_{2\varepsilon}$ tend to infinity when ε tends to zero. Hence the somewhat akward definition (or some definition like it) of $Z_{2\varepsilon}$ is necessary.

Proof of lemma 4.1

We will use the martingale-central-limit theorem as stated in Andersen&Gill[1982] (Theorem I.2.).

It is straightforward to verify that we may write

$$Z_1^n(t) = \sqrt{n} \int_0^t \Delta(Y_1(s)>0)\ Y_1(s)^{-1}\ dM_{13}(s)$$

(where we interpret 0/0=0). Hence $Z_1^n(t)$ is an integral of a bounded and predictable process adapted to $\{\mathcal{F}_t\}$ with respect to the martingale of $N_{13}(t)$ in $\{\mathcal{F}_t\}$

Also we may write

$$Z_{2\varepsilon}^n(t) = 1/\sqrt{n}\ \{ \int_0^t \Delta(s>\varepsilon)\ J_{13}(s)\ [S(s)\ (1-S(s))^{-1} dM_{23}(s)\ -\ dM_{01}(s)]$$

$$+ \int_0^t \Delta(s>\varepsilon) J_{13}(s)(Y_0(s)+Y_1(s))[S(s)\ (1-S(s))^{-1}(1-S(s))-S(s)]\gamma(s)ds\ \}$$

where the last integral (always) cancels out. Hence $Z_{2\varepsilon}^n(t)$ is the sum of two integrals of predictable and bounded processes with respect to the martingales $M_{01}(t)$ and $M_{23}(t)$ which are the martingales of $N_{01}(t)$ and $N_{23}(t)$ with respect to $\{\mathcal{F}_t\}$.

Hence the conclusion follows from

$$\langle Z_1^n\rangle(t) = \int_0^t n\ \Delta(Y_1(s)>0)\ Y_1(s)^{-1}\ \alpha(s)\ ds \ \xrightarrow{P}\ \int_0^t y_1(s)^{-1}\ \alpha(s)\ ds$$

and

$$\langle Z_{2\varepsilon}^n\rangle(t) = \int_0^t \Delta(s>\varepsilon)\ J_{13}(s)\ S(s)\ (1-S(s))\ n^{-1}\ (Y_0(s)+Y_2(s))\ \gamma(s)\ ds$$

$$\xrightarrow{P} (1-\lambda) \int_0^t \Delta(s>\varepsilon)\ S(s)\ (1-S(s))^{-1}\ g(s)\ ds$$

on D[0,1] which is straightforward consequences of Lemma A.1.

We have here verified condition I.3 in the Andersen&Gill-theorem. Condition I.4 is trivial since the predictable processes we are integrating with respect to are bounded.

Lemma 4.2

Let $V_{n\varepsilon}(t) = Z^n_{1\varepsilon}(t) + V_{n2\varepsilon}(t) + V_{n3\varepsilon}(t)$

where

$$V_{n2\varepsilon}(t) = \int_\varepsilon^t [Z^n_{2\varepsilon}(1) - Z^n_{2\varepsilon}(s)]\ y^*(s)^{-1}\ \alpha(s)\ ds$$

and

$$V_{n3\varepsilon}(t) = -(1-\lambda)\int_\varepsilon^t \int_s^1 S(r)\ (1-S(r))^{-1}\ Z^n_1(r)\ g(r)dr\ y^*(s)^{-1}\ \alpha(s)\ ds$$

Then for all $\varepsilon>0$ we have $\sqrt{n}\ (\beta^*_\varepsilon(t)-\beta_\varepsilon(t)) - V_{n\varepsilon}(t) \overset{P}{\to} 0$ on $D[\varepsilon,1]$.

Proof of lemma 4.2

We may write

$\sqrt{n}\ (\beta^*_\varepsilon(t)-\beta_\varepsilon(t)) = \sqrt{n}\ (\beta^*_\varepsilon(t)-\hat\beta_\varepsilon(t)) + \sqrt{n}\ (\beta_\varepsilon(t)-\beta_\varepsilon(t))$ and

it is easy to prove that $\sqrt{n}\ (\hat\beta_\varepsilon(t)-\beta_\varepsilon(t)) - Z^n_{1\varepsilon}(t) \overset{P}{\to} 0$ on $D[\varepsilon,1]$

for $\varepsilon>0$. Hence it will suffice to show

$$\sqrt{n}\ (\beta^*_\varepsilon(t)-\hat\beta_\varepsilon(t)) - V_{n2\varepsilon}(t) - V_{n3\varepsilon}(t) \overset{P}{\to} 0 \text{ on } D[\varepsilon,1] \text{ for } \varepsilon>0. \tag{4.7}$$

Let

$$V^*_{2n\varepsilon}(t) = \int_\varepsilon^t [Z^n_{2\varepsilon}(1) - Z^n_{2\varepsilon}(s)]\ (n/Y^*(s))\ Y_1(s)^{-1}\ dN_{13}(s)$$

and

$V^*_{n3\varepsilon}(t)=$

$$\sqrt{n}\int_\varepsilon^t\int_s^1(\hat S(r)-S(r))(1-\hat S(r))^{-1}\ J_{13}(r)(1-S(r))^{-1}dN_{23}(r)\ Y^*(s)^{-1}\ Y_1(s)\ dN_{13}(s)$$

We will first demonstrate that

$$\sqrt{n}(\beta^*_\varepsilon(t)-\beta_\varepsilon(t)) = V^*_{n2\varepsilon}(t) + V^*_{n3\varepsilon}(t) \tag{4.8}$$

and afterwards that on $D[\varepsilon,1]$ for $\varepsilon>0$

$$V^*_{n2\varepsilon}(t) - V^*_{n2\varepsilon}(t) \overset{P}{\to} 0 \tag{4.9}$$

and

$$V^*_{n3\varepsilon}(t) - V^*_{n3\varepsilon}(t) \overset{P}{\to} 0 \tag{4.10}$$

From (4.8)-(4.10) we obviously get (4.7) and hence the lemma will follow.

By (2.8) and (2.12) (def av $\hat\beta$ og β^*) we get

$$\sqrt{n}(\beta^*_\varepsilon(t)-\hat\beta_\varepsilon(t)) = \sqrt{n}\int_\varepsilon^t [\ Y^{**}(s)-Y^*(s)\ ]\ [Y^*(s)Y_1(s)]^{-1}\ dN_{13}(s) \tag{4.11}$$

and by expanding

$$Y^{**}(s)-Y^{*}(s) = \int_s^1 \hat{S}(r)\ (1-\hat{S}(r))^{-1} J_{13}(r)dN_{23}(r) - \int_s^1 J_{13}(r)dN_{01}(r)$$

$$= \int_s^1 [\hat{S}(r)(1-\hat{S}(r))^{-1} - S(r)(1-S(r))^{-1}]\ J_{13}(r)\ dN_{23}(r)$$

$$+[\ \int_s^1 S(r)(1-S(r))^{-1} J_{13}(r)\ dN_{23}(r) - \int_s^1 J_{13}(r)\ dN_{01}(r)\ ]$$

this give

$$Y^{**}(s)-Y^{*}(s) = \int_s^1 (\hat{S}(r)-S(r))(1-\hat{S}(r))^{-1}(1-S(r))^{-1} J_{13}(r)dN_{23}(r)$$
$$+ \sqrt{n}\ [\ Z^n_{2\varepsilon}(1) - Z^n_{2\varepsilon}(s)\ ] \tag{4.12}$$

By inserting (4.12) into (4.11) we get (4.8).

Next we will demonstrate (4.9) is a consequence of lemma A.5. Too see this note that

$$V_{n2\varepsilon}(t)-V^{*}_{n2\varepsilon}(t) = \int_\varepsilon^t [Z^n_{2\varepsilon}(1)-Z^n_{2\varepsilon}(s)]\ y^{*}(s)^{-1}\alpha(s)ds$$
$$- \int_\varepsilon^t [Z^n_{2\varepsilon}(1)-Z^n_{2\varepsilon}(s)]\ (n/Y^{*}(s))\ Y_1(s)\ dN_{13}(s)$$

But

$$\int_\varepsilon^t (n/Y^{*}(s))\ Y_1(s)^{-1}\ dN_{13}(s) \xrightarrow{P} \int_\varepsilon^t y^{*}(s)^{-1}\ \alpha(s)\ ds \quad \text{on } D[\varepsilon,1]$$

while

$$Z^n_{2\varepsilon}(1) - Z^n_{23}(s) \xrightarrow{D} Z_{2\varepsilon}(1) - Z_{2\varepsilon}(s) \quad \text{on } D[\varepsilon,1]$$

which is a process on $C[\varepsilon,1]$ with probabilty 1. Hence (A.6) applies and

$$V_{2n\varepsilon}(t)-V^{*}_{2n\varepsilon}(t) \xrightarrow{P} 0 \quad \text{on } D[\varepsilon,1].$$

To prove (4.10) we will write $V_{n3\varepsilon}(t) - V^{*}_{n3\varepsilon}(t)$

$$=\int_\varepsilon^t \int_s^1 \frac{-S(r)}{1-S(r)}\ Z^n_1(r)\ [(1-\lambda)g(r)dr - \frac{J_{13}(r)dN_{23}(r)}{n\ (1-S(r)}\ ]\ y^{*}(s)^{-1}\alpha(s)ds$$
$$+\int_\varepsilon^t \int_s^1 \frac{-S(r)}{1-S(r)}\ Z^n_1(r)\ J_{13}(r)\ \frac{dN_{23}(r)}{n\ (1-S(r)}\ [\ y^{*}(s)^{-1}\alpha(s)ds - \frac{n\ dN_{01}(s)}{Y^{*}(s)Y_1(s)}\ ]$$
$$+\int_\varepsilon^t \int_s^1 -S(r)\ Z^n_1(r)\ [\ \frac{J_{13}(r)}{1-S(r)} - \frac{J_{13}(r)}{1-\hat{S}(r)}\ ]\ \frac{dN_{23}(r)}{n\ (1-S(r)}\ \frac{n\ dN_{01}(s)}{Y^{*}(s)Y_1(s)}$$
$$+\int_\varepsilon^t \int_s^1 [-S(r)\ Z^n_1(r) - \sqrt{n}(\hat{S}(r)-S(r))]\ \frac{J_{13}(r)}{1-\hat{S}(r)}\ \frac{dN_{23}(r)}{n\ (1-S(r)}\ \frac{n\ dN_{01}(s)}{Y^{*}(s)Y_1(s)} \tag{4.13}$$

We will show that each of the four terms in (4.13) will tend to zero in probability on $D[\varepsilon,1]$.

The argument that the inner integral in the first term will "vanish" is analogue to the demonstration of (4.9). Hence using lemma A.1 the first term "disappears".

The second term disappears since by (A.5) in lemma A.5 the inner integral will tend to a process on $D[\varepsilon,1]$ which is gaussian with continous covariance function hence on $C[\varepsilon,1]$ with probability 1. Using lemma A.5 (A.6) on the outer integral the conclusion follows.

Since $\hat{S}(s) \xrightarrow{P} S(s)$ on $D[0,1]$, it will also hold on $D[\varepsilon,1]$ that $J_{13}(s)\,[\,(1-S(s))^{-1} - (1-\hat{S}(s))\,]^{-1} \xrightarrow{P} 0$. Using lemma A.1 twice show that the third term converges (in probability) to zero on $D[\varepsilon,1]$.

From Gill[1980] theorem 3.2.1(iv) we get

$$[-S(r)Z_1^n(r) - \sqrt{n}(\hat{S}(r)-S(r))] = -S(r)\int_0^r (1-\hat{S}(s-)/S(s))\, dZ_1^n(s) \qquad (4.14$$
$$-\sqrt{n}\; S(r)\int_0^r \hat{S}(s-)/S(s)\; \Delta(Y_1(s)>0)\; \alpha(s)ds$$

Using Theorem I.2. in Andersen&Gill[1982] it is easy to show that the first term in (4.14) converges weakly on $D[0,1]$ to a gaussian process with variance function equal to zero, hence it converges to zero in probability on $D[0,1]$. In the last term of (4.13) the right-handside of (4.14) is multiplied with $J_{13}(r)(1-\hat{S}(r))^{-1}$ which converges to $(1-S(r))^{-1}$ in probability on $D[\varepsilon,1]$. Arguing similarily as with the third term we see that also the fourth term will vanish on $D[\varepsilon,1]$.

Consequently $V_{n3\varepsilon}(t) - V^*_{n3\varepsilon}(t) \xrightarrow{P} 0$ $D[\varepsilon,1]$ and lemma 4.2 follows.

<u>Proof of Proposition 4.1</u>

Let

$$\breve{V}_\varepsilon(u) = Z_1(u) - Z_1(\varepsilon) + \int_\varepsilon^u [\,Z_{2\varepsilon}(1) - Z_{2\varepsilon}(t)\,]\; y^*(t)^{-1}\; \alpha(t)\; dt \qquad (4.15)$$
$$-\; (1-\lambda)\int_\varepsilon^u \int_t^1 S(s)\;(1-S(s))^{-1}\; Z_1(s)\; g(s)\; ds\; y^*(t)^{-1}\; \alpha(t)\; dt$$

Then it is easy to see that the functional Φ: $D[0,1]^2 \to D[\varepsilon,1]$ given by $\Phi(Z_1, Z_{2\varepsilon})(u) = \breve{V}_\varepsilon(u)$ is continous. Hence

$$V_{n\varepsilon}(u) = \Phi(Z_1^n, Z_{2\varepsilon}^n)(u) \xrightarrow{D} \Phi(Z_1, Z_{2\varepsilon})(u) = \breve{V}_\varepsilon(u)$$

on $D[\varepsilon,1]$.

Since Z_1 and $Z_{2\varepsilon}$ are gaussian with expectation 0 it is also easy to prove that $\breve{V}_\varepsilon$ is gaussian with expectation 0. To show that $\breve{V}_\varepsilon \overset{D}{=} V_\varepsilon$ defined in proposition 4.1 (and hence to complete the proof) it will suffice to show that the covariance function of $\tilde{V}_\varepsilon$ is the same as the covariance function of V_ε (given by (4.4)). This will follow when we show that

$$\tilde{V}_\varepsilon(u) = \int_0^1 H^u_{1\varepsilon}(t)\, dZ_1(t) + \int_\varepsilon^1 H^u_{2\varepsilon}(t)\, dZ_{2\varepsilon}(t) \tag{4.16}$$

since

$$E\,\tilde{V}_\varepsilon(u)\,\tilde{V}_\varepsilon(t) = E\int_0^1 H^u_{1\varepsilon}(s)\, H^t_{1\varepsilon}(s)\, d\langle Z_1, Z_1\rangle(s)$$
$$+ E\int_\varepsilon^1 H^u_{2\varepsilon}(s)\, H^t_{2\varepsilon}(s)\, d\langle Z_{2\varepsilon}, Z_{2\varepsilon}\rangle(s)$$
$$= c_\varepsilon(u,t)$$

The last equation follows because

$$d\langle Z_1, Z_1\rangle(s) = y(s)^{-1}\,\alpha(s)\,ds$$

and $d\langle Z_{2\varepsilon}, Z_{2\varepsilon}\rangle(s) = (1-\lambda)S(s)(1-S(s))^{-1}g(s)ds$

The formula (4.16) is derived by partial integrations. First demonstrate how the second term in (4.16) arises.

$$\int_\varepsilon^u [Z_{2\varepsilon}(1)-Z_{2\varepsilon}(t)]\, y^*(t)^{-1}\alpha(t)dt = \int_\varepsilon^u \int_\varepsilon^t y^*(s)^{-1}\alpha(s)ds\, [Z_{2\varepsilon}(1)-Z_{2\varepsilon}(t)]$$
$$+ \int_\varepsilon^u \int_\varepsilon^t y^*(s)^{-1}\alpha(s)ds\, dZ_{2\varepsilon}(t)$$

$$= \int_\varepsilon^1 \int_\varepsilon^{t\wedge u} y^*(s)^{-1}\alpha(s)ds\, dZ_{2\varepsilon}(t) = \int_\varepsilon^1 H^u_{2\varepsilon}(t)\, dZ_{2\varepsilon}(t) \tag{4.17}$$

Similarily

$$\int_s^1 S(r)\,(1-S(r))^{-1}\, g(r)\, Z_1(r)\, dr = \int_s^1 -\int_t^1 S(r)\,(1-S(r))^{-1}\, g(r)\, dr\, Z_1(t)$$
$$+ \int_s^1 \int_t^1 S(r)\,(1-S(r))\, g(r)\, dr\, dZ_1(t)$$
$$= \int_0^1 \int_{s\vee t}^1 S(r)\,(1-S(r))\, g(r)\, dr\, dZ_1(t)$$

Hence by inserting this into the third term in (4.15) we see

$$Z_1(u)-Z_1(\varepsilon) - (1-\lambda)\int_\varepsilon^u \int_t^1 S(s)\,(1-S(s))^{-1}\, Z_1(s)\, g(s)\, ds\, y^*(t)^{-1}\,\alpha(t)\, dt$$
$$= \int_0^1 H^u_{1\varepsilon}(t)\, dZ_1(t) \tag{4.18}$$

From (4.17) and (4.18) we see that (4.15) equals (4.16). Thus the proof of proposition 4.1 is complete.

It is of great interest to esimate the covariance-function $c(u,t)$. We will here suggest a natural choice. This choice is constructed the following way: All terms $\alpha(t)dt$ in $c(u,t)$ are substituted by the terms $Y^*(s)^{-1}\, dN^*(s)$ and correspondingly all terms $(1-\lambda)\, g(s)\, ds$ are replaced by $n^{-1}(dN_{01}(s)+dN_{23}(s))$. Also for $S(t)$ in $c(u,t)$ we use $S^*(t)$, for $y_1(t)$ we use $Y_1(t)/n$ and for $y^*(t)$ we use $Y^*(t)/n$. Let the estimator that arises be denoted $C^*(u,t)$ and the corresponding estimator for $c_\varepsilon(u,t)$ be denoted $C^*_\varepsilon(u,t)$. In Samuelsen[1984] it is shown that $\sup_{0\leq u\leq 1,} | C^*_\varepsilon(u,u) - c_\varepsilon(u,u) | \xrightarrow{P} 0$ as $n \to \infty$. (Actually some algebra is neccesary to show that this is what is shown.) To prove that also $C^*_\varepsilon(u,t)$ and $C^*(u,t)$ (for all t and u) are consistent are fairly easy modifications of this proof.

Section 5. Convergence of the "One-Step P-L-estimator".

Up to now we have solely focused upon the asymptotic properties of $\beta^*(t)$. In this section we will give the corresponding results for the estimator $S^*(t)$ (given by (2.11)). We will also indicate how the results may be proved.

Theorem 5.1

$$\sup_{0\leq t\leq 1} |\ S^*(t) - S(t)\ | \xrightarrow{P} 0 \quad \text{as} \quad n \to \infty$$

Theorem 5.2

Let V(t) be as given in Theorem 4.1. Then on D[0,1]

$$\sqrt{n}\ (\ S^*(t) - S(t)\ ) \xrightarrow{D} -S(t)\ V(t)$$

Thus $S^*(t)$ is consistent and $\sqrt{n}(S^*(t) - S(t))$ converges weakly to a gaussian process. This hold because $S^*(t)$ is approximatly equal $\exp(-\beta^*(t))$. Let

$$R_n^*(t) = S^*(t)/\exp(-\beta^*(t)) = \exp(\ \int_0^t \ln(1-Y^*(s)^{-1}dN^*(s))\ \ \exp(\beta^*(t))$$

By a Taylor expansion it is seen

$$R_n^*(t) = \exp\{\ \int_0^t\ \sum_{k=2}^{\infty} -\frac{1}{k}\ [dN^*(s)/Y^*(s)]^k\ \}$$

Then we may show that both

$$\sup_{0\leq t\leq 1} |\ R_n^*(t) - 1\ | \xrightarrow{P} 0 \tag{5.1}$$

and

$$\sup_{0\leq t\leq 1} \sqrt{n}\ |\ R_n^*(t) - 1\ | \xrightarrow{P} 0 \tag{5.2}$$

hold. Hence we get

$$\sup_{0\leq t\leq 1} |\ S^*(t) - S(t)\ | \leq \sup_{0\leq t\leq 1} |\ \exp(-\beta^*(t)) - \exp(-\beta(t))\ | + \sup_{0\leq t\leq 1} \exp(-\beta^*(t))\ |\ R^*(t) - 1\ |$$

The first term disappear due to theorem 4.1, while the second term will vanish because of (5.2) and theorem 5.1 follows.

Furthermore

$$\sqrt{n}(S^*(t) - S(t)) = \sqrt{n}\ \{\ \exp(-\beta^*(t)) - \exp(-\beta(t))\ \} + \sqrt{n}\ \exp(-\beta^*(t))\ \{\ R_n^*(t) - 1\ \}$$

where the last term tends to zero uniformly on D[0,1] in probabilty. Also the first term will be asymptotically equivalent to

$$-S(t)\ \sqrt{n}\ \{\ \beta^*(t) - \beta(t)\ \}$$

which will have the limit process given in theorem 5.2.

Section 6. Simulation study.

Allthough we have proved some results concerning the asymptotic behaviour of β^*, many questions regarding this estimator remains completely unanswered. For instance it would be nice to have some idea of how much (if anything) we gain by using β^* instead of $\hat{\beta}$. Further we would like to know something about the loss of not iterating to convergence(i.e. to use β^* instead of the Turnbull-estimator β^{∞}.) In addition it is of interest to study the small-sample properties (as the bias and the fit to the normal distribution) of the estimators.

To tentatively answer these questions we have performed a simulation experiment. The model that was simulated was a markov-process on the state space in figur 1.1 with all transition-intensities equal to 1. Hence T_i, $L_i | L_i > 0$ and $U_i - L_i$ are all exponentially distibuted with expectation 1. Especially the cumulative intensity $\beta(t)=t$.

The model was simulated for varying samle-size n=30,100,300 and varying $\lambda=Pr(L_i>0)=0.1,0.3,0.5,0.8$. In each simulation (n,λ) we registrated $\hat{\beta}(t),\beta^*(t),\beta^{\infty}(t),C_1^*(t,t)$ and $C^*(t,t)$ where the last two quantities are estimators of the variance of $\sqrt{n}(\hat{\beta}(t)-\beta(t))$ and $\sqrt{n}(\beta^*(t)-\beta(t))$ (respectively). In section 4 it is explained how C^* is constructed. C_1^* is constructed from the same principles.
Each simulation was performed 200 times and the following summary statistics was computed for $t=0.1,0.2,\ldots\ldots,2.0$.

$\hat{\beta}.(t)$=average of the 200 estimates of $\beta(t)$

$\beta^*.(t)$=average of the $\beta^*(t)$

$\beta^{\infty}.(t)$=average of the Turnbull-estimators $\beta^{\infty}(t)$

$C_1^*.(t)$=average of the $C_1^*(t,t)$=average of an estimates of the limiting-distribution of $\sqrt{n}(\hat{\beta}(t)-\beta(t))$

$C^*.(t)$=average of the $C^*(t,t)$

$C.^{**}(t)$=empirical variance of the $\beta^*(t)$, normalized

$C.^{\infty}(t)$=empirical variance of the $\beta^{\infty}(t)$, normalized

$H^*(t)$=histogram of the standarized $\beta^*(t)$ (i.e. $H^*(t)$ counts the number of the $\sqrt{n}(\beta^*(t)-\beta(t))/(C^*.(t))^{1/2}$ in the intervals $[-0.25,0.25],\langle 0.25,0.75],\langle -0.25,-0.75],\ldots\ldots$)

First we look at the bias of β^* and β^∞. In Figure 6.1 we have drawn $\hat{\beta}.(t)-\beta(t)$, $\beta^*.(t)-\beta(t)$ and $\beta^\infty.(t)-\beta(t)$ for $(n,\lambda)=(100,0.1)$. The picture displayed here, that both $\beta^*(t)$ and $\beta^\infty(t)$ appears to underestimate $\beta(t)$ with the bias of $\beta^\infty(t)$ slightly worse than (the bias of) $\beta^*(t)$, seems to be confirmed also for ather choiches of (n,λ).

However, the bias seems to disappear quite fast, as is indicated by Figure 6.2 which show $\beta^*.(t)-\beta(t)$ for $\lambda=0.1$ and $n=30,100,300$.

Next we turn to the question of whether $C^*(t,t)$ does estimate the variance of $\sqrt{n}\,(\beta^*(t)-\beta(t))$. A failure to do so may be due to two factors, i.e. $C^*(t,t)$ is a biased estimator of the asymptotic variance of $\sqrt{n}\,(\beta^*(t)-\beta(t))$ or the asymptotic variance does not give a good approximation to the true variance. In Figur 6.3 $C^*.(t)$ and $C^{**}.(t)$ is plotted for $(n,\lambda)=(300,0.1)$; $C^*.(t)$ being the smooth curve.

For small samples, however, the fit may not at all be good. As an example see Figure 6.4 which give the corresponding curves for $(n,\lambda)=(30,0.1)$. A closer inspection indicates that the difference between $C^*.(t)$ and $C^{**}.(t)$ for small n seems to be a consequence of both that $C^*(t,t)$ overestimates the asymptotic variance and that the asymptotic variance is greater than the true variance.

Furthermore we look at the (anticipated) gain in using $\beta^*(t)$ instead of $\hat{\beta}(t)$. In Figure 6.5 $C^*_1.(t)/C^*.(t)$ are drawn for $n=300$ and $\lambda=0.1$, 0.3, 0.5 and 0.8 . It seems like a considerable degree of left-censoring cooresponds to a considerable gain.

It is somewhat unsatifactory not having any analytical results for $\beta^\infty(t)$. In this respect Figure 6.6 give some relief. It show $C^{**}.(t)/C^\infty.(t)$ for $n=100$ and $\lambda=0.1,0.3,0.5$. This figure indicates a loss in efficiency up to 25% (for $\lambda=0.1$) of using $\beta^*(t)$ instead of $\beta^\infty(t)$. However for $n=30$ and $n=300$ the estimated efficiency function $C^{**}.(t)/C^\infty.(t)$ lies (even) closer to 1.

Finally we take a look at the normal-aproximation of $\sqrt{n}(\beta^*(t)-\beta(t))$. In Figure 6.7 we have drawn $H^*(t)$ for $(n,\lambda,t) = (30,0.1,2.0)$, (300,0.1,2.0) and (100,0.1,1.0) (in that order). These histograms indicate the general picture in the sense that in most cases the histograms are onetopped, somewhat left-skiew with heavier tails than the normaldistribution. It is not obvious from the histograms that the approximation to the normaldistribution improves as n gets larger.

Figure 6.1

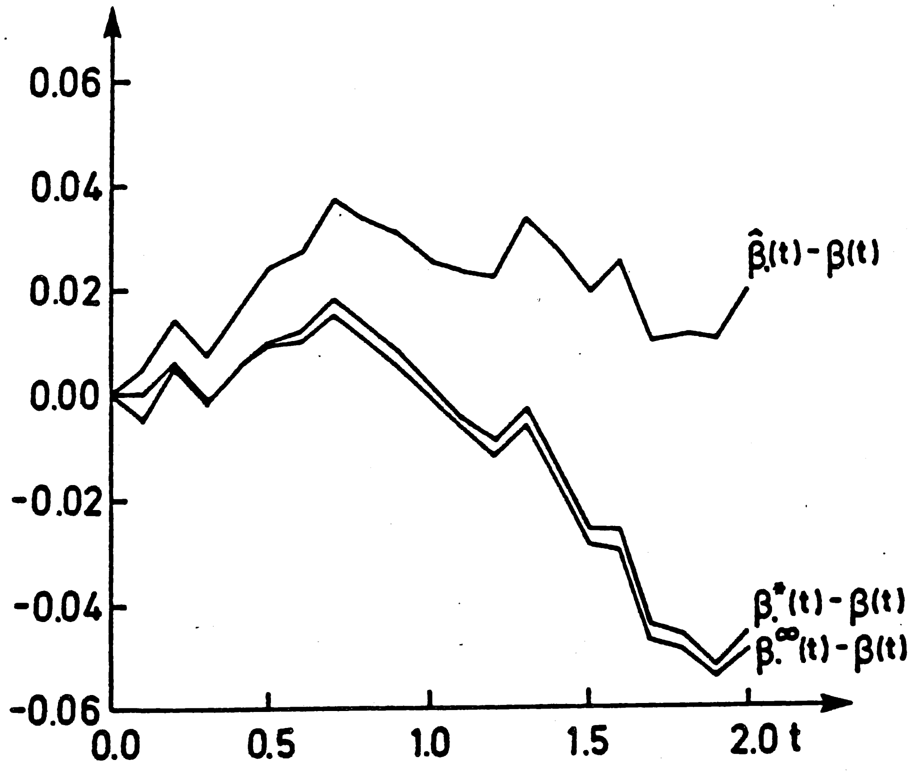


Figure 6.2

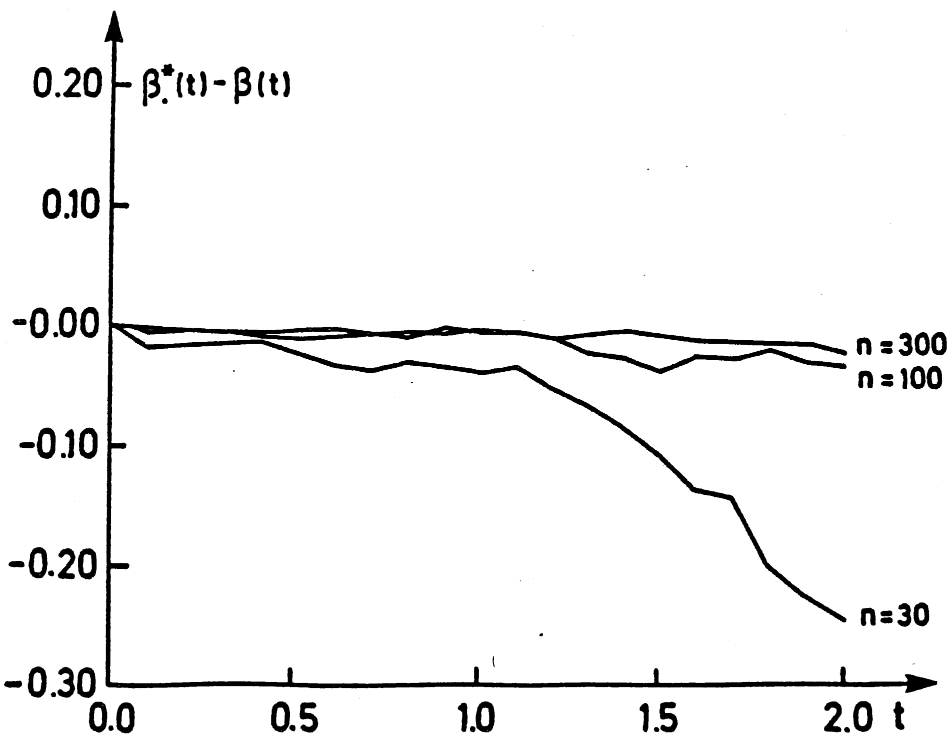


Figure 6.3

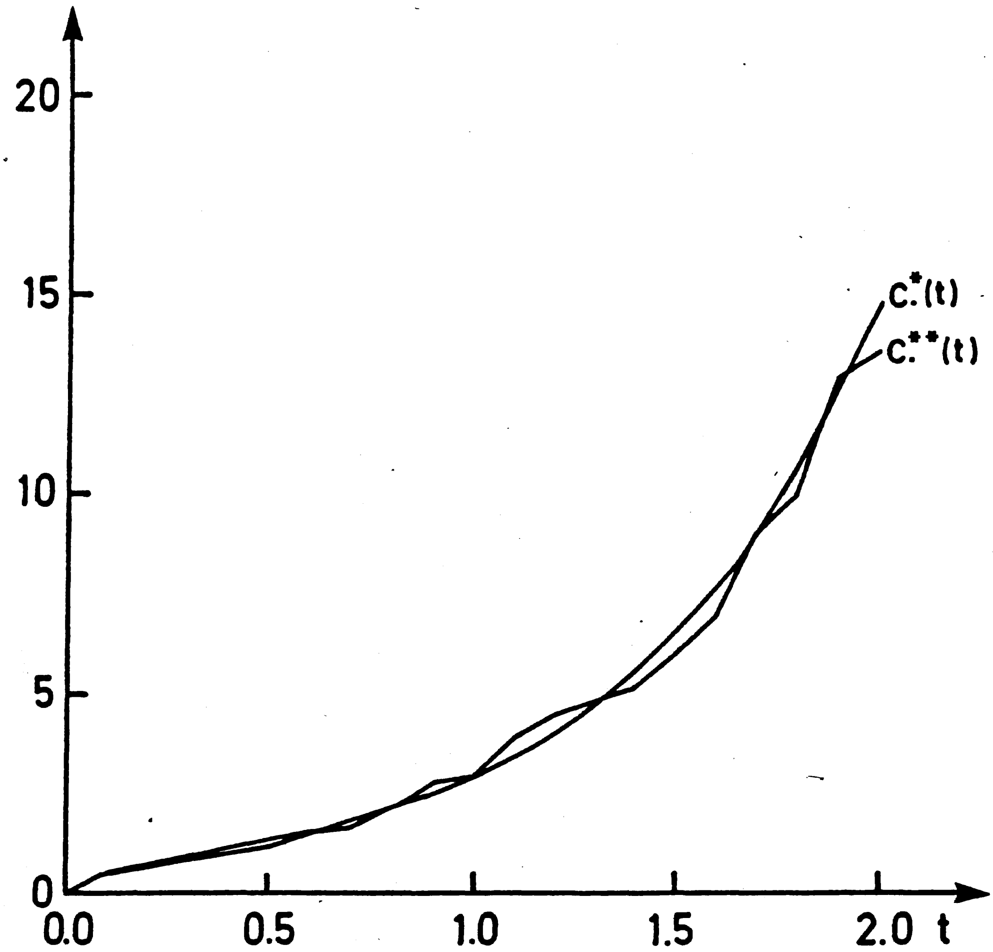


Figure 6.4

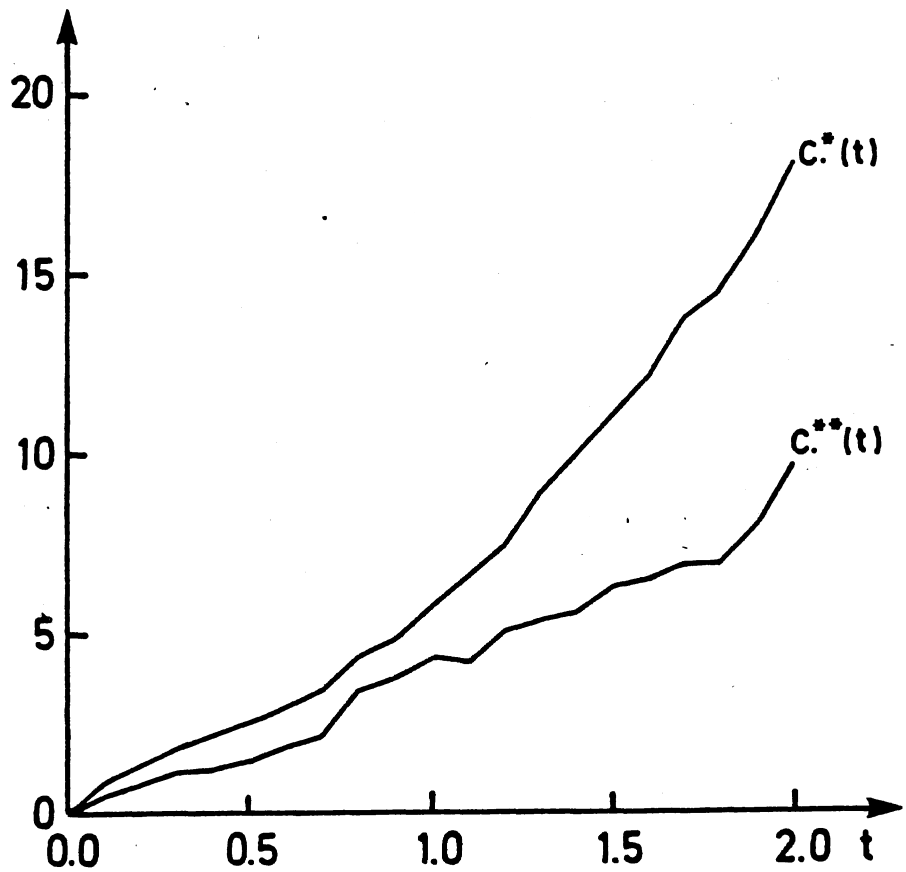


Figure 6.5

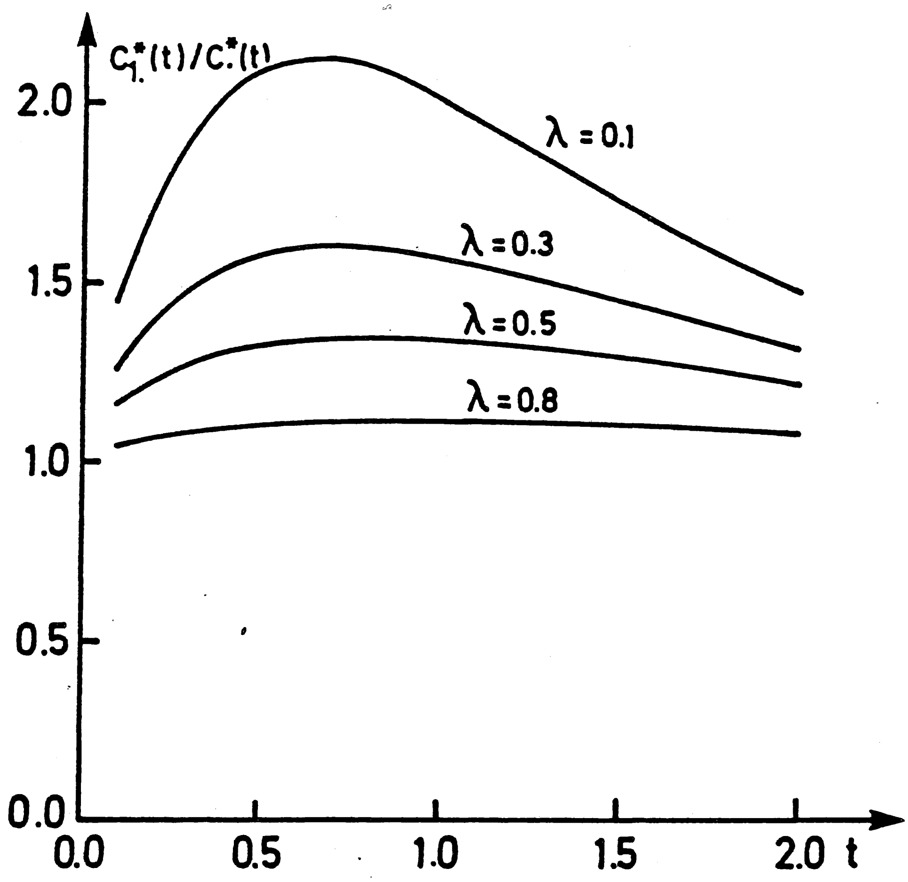


Figure 6.6

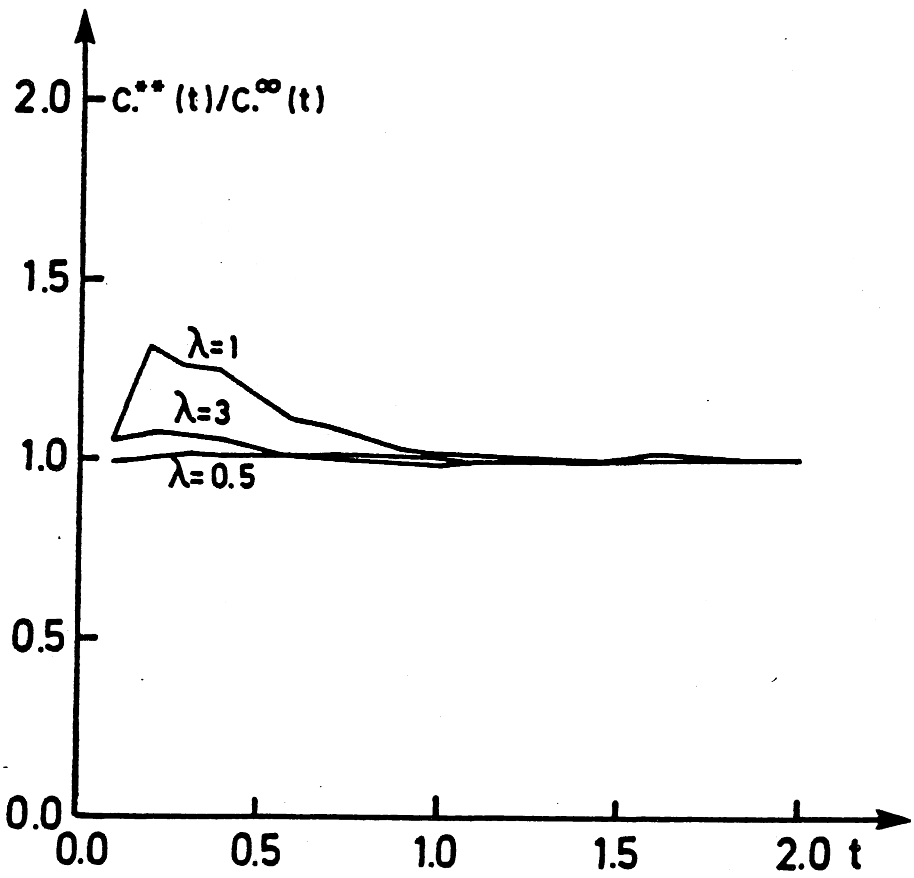
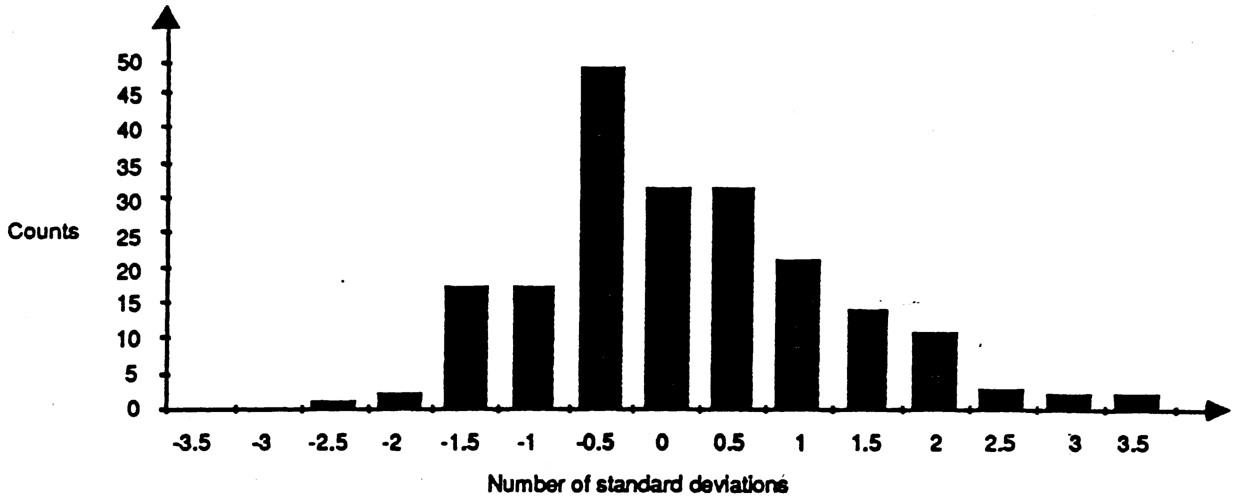
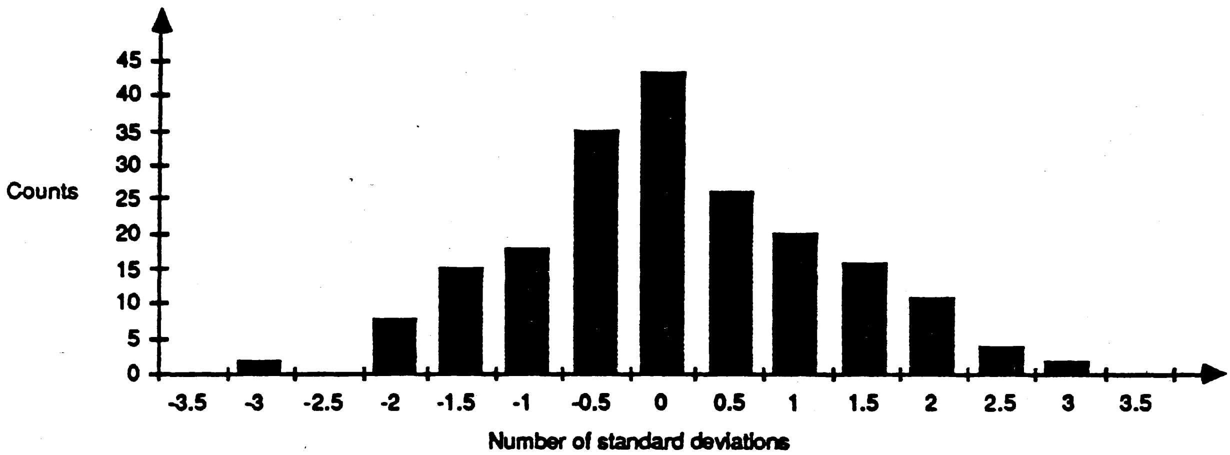
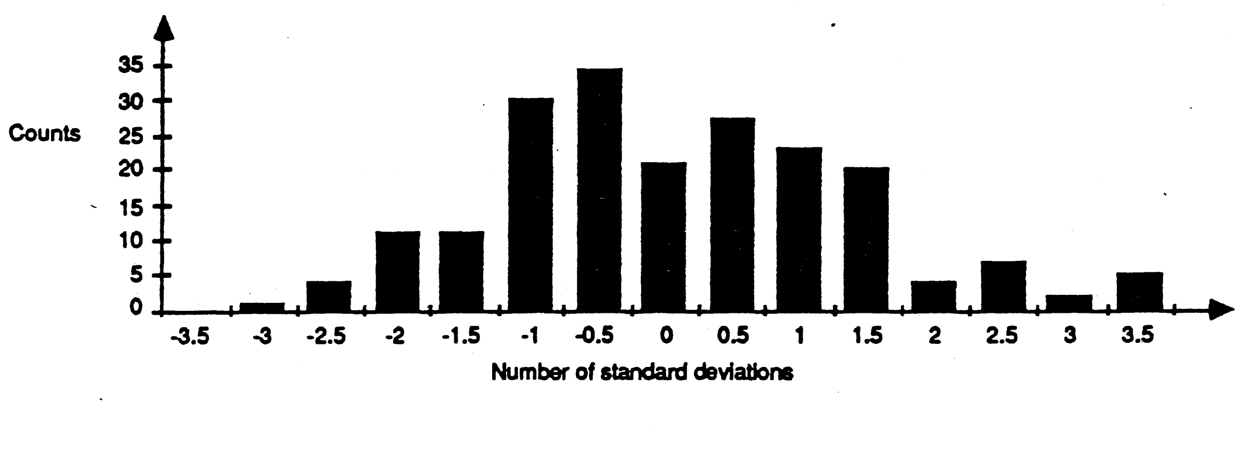


Figure 6.7


50
45
40
35
30
Counts
25
20
15
10
5
0
-3.5
-3
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
3
3.5
Number of standard deviations


45
40
35
30
Counts
25
20
15
10
5
0
-3.5
-3
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
3
3.5
Number of standard deviations


35
30
Counts
25
20
15
10
5
0
-3.5
-3
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
3
3.5
Number of standard deviations

Section 7. Discussion.

First we want to comment upon some of the assumptions made in Section 1.

Of the distributional assumptions 1), 2) and 3) are clearly necessary. If they should be left out or modified the posed problem would also be modified. For instance if 3) was left out this would correspond to allowing truncated observations (with $T_i \leq 1$) and the estimation procedure would quite different (see e.g. Turnbull[1976].) In order to determine $\beta(t)$ on all of [0,1] also assumption 5) is necessary, however if it were sufficient to determine only $\beta(t)-\beta(\varepsilon)$ for some $\varepsilon>0$ and $t>\varepsilon$ a weaker version of 5) might be sufficient. Assumption 4) is almost abundant in view of 5) and is mostly a convenient way of introducing notation.

Regarding assumption 6) , we conjecture that most result will still hold also for general distributions (e.g. for discreet or general mixtures of discreet and contionous) but that the expression for the covariance-function will be more comnplicated.

Next we turn to the observational assumptions. These only differ from the assumptions in Turnbull[1974] in that we demand L_i to be observed. The limit of the iteration (i.e. the Turnbull-estimator) is not affected of this assumption and hence neither is its properties. Allthough we have proved no such thing, it seems reasonable that (at least the large-sampel properties of) the Turnbull-estimator is not inferior to the "one-step-estimator". This is also confirmed by the simulation experiment.

The variance-function estimator of $\beta^*(t)$ may be modified (by among other things substituting $S^*(t)$ by $S^{**}(t)$) so that it is possible to compute even if L_i is not observed. Hence it is possible to construct an (over-)estimator of $\beta^{**}(t)$.

If there is a vast amount of left-censoring present one might fear that there is a great loss in not iterating furhter. But it should be straightforward (even if it is quite tedious) to derive the asymptotic properties of the "second-iteration-estimator" or of any "finite-iteration-estimator". For instance in Samuelsen[1984] it is proved that the "n-th iteration estimator" is consitent for any finite n. The complexity of the covariance-function however will increase with the number of iterations.

Acknowledgments: This paper consists mainly of the authors graduate thesis. I would like to express my thanks to my supervisor Odd Aalen for introducing me to the problem and for support throughout the work. Also I would like to thank Ørnulf Borgan for many helpful suggestions. Figur 6.7 was made by the kind help of Arne Bang Huseby.

Appendix A. Lemmas omitted from the main text.

Lemma A.1

Suppose $\{ F_n ; n=1,2,\ldots \}$ and $\{ g_n ; n=1,2,\ldots \}$ are sequences of stochastic functions on [0,1] and F_0 and g_0 are deterministic contionous functions on [0,1] which satisfy

1) F_n are nondecreasing, rightcontionous with $F_n(0) = 0$ and $F_n(1) \leq M < \infty$ with probability 1 for n=0,1,2,...
2) $\sup_{t\in[0,1]} | F_n(t) - F_0(t) | \overset{a.s}{\to} 0$ as $n \to \infty$
3) $\sup_{t\in[0,1]} | g_n(t) - g_0(t) | \overset{a.s}{\to} 0$ as $n \to \infty$

Then

$$\sup_{t\in[0,1]} | \int_0^t g_n(s)\, dF_n(s) - \int_0^t g_0(s)\, dF_0(s) | \overset{a.s}{\to} 0 \quad \text{as} \quad n \to \infty .$$

If everywhere in the theorem we replace a.s. with P the result is still valid.

lemma A.2

Let 2) and 3) in lemma A.1 be replaced by

4) $\sup_{t\in[0,1]} | F_n(t) - F_0(t) | \overset{P}{\to} 0$ as $n \to \infty$
5) For all $\varepsilon > 0$ we have
$$\sup_{t\in[\varepsilon,1]} | g_n(t) - g_0(t) | \overset{P}{\to} 0 \quad \text{as} \quad n \to \infty$$
6) There exist a sequnce of stochastic function A_n such that for all $\varepsilon>0$ $P(\int_0^\varepsilon | g_n(s) | \, dF_n(s) \leq A_n(\varepsilon)) \to 1$ and with $\lim_n EA_n(\varepsilon) \to 0$ as $\varepsilon \to 0$.

Then we have

$$\sup_{t\in[0,1]} | \int_0^t g_n(s)\, dF_n(s) - \int_0^t g_0(s)\, dF_0(s) | \overset{P}{\to} 0$$

Proof of Lemmas 3.1 and 3.2

Lemma A.1 is a straightforward consequence of lemma 6.1 in Aalen[1976]. Restating lemma A.2 we have that for all $\gamma>0$ and $\delta>0$ there exist an N such that for n>N

$$P(\sup_{t\in[0,1]} | \int_0^t g_n\, dF_n - \int_0^t g_0\, dF_0 | > \delta) < \gamma$$

But

$$\sup_{t\in[0,1]} \left| \int_0^t g_n \, dF_n - \int_0^t g_0 \, dF_0 \right| \le \int_0^\varepsilon |g_n| dF_n + \int_0^\varepsilon |g_0| dF_0$$

$$+ \sup_{t\in[\varepsilon,1]} \left| \int_\varepsilon^t g_n \, dF_n - \int_\varepsilon^t g_0 \, dF_0 \right|$$

By lemma 6.1 in Aalen[1976] the last term will for all $\varepsilon>0$ have probability less than $\gamma/2$ of being greater than $\delta/3$ if $n>N=N(\varepsilon)$. Further since g_0 is integrabel with respect to F_0 we have $\int_0^\varepsilon |g_0| dF_0 < \delta/3$ for sufficently small ε. Finally notice that

$$P\left[\int_0^\varepsilon |g_n| dF_n > \delta/3 \right] \le P\left[A_n(\varepsilon) < \int_0^\varepsilon |g_n| dF_n \right] + P\left[A_n(\varepsilon) > \delta/3 \right]$$

$$\le P\left[A_n(\varepsilon) < \int_0^\varepsilon |g_n| dF_n \right] + 3 \, EA_n(\varepsilon)/\delta$$

By assumption we can find n sufficiently large and ε sufficiently small so that the right-hand-side of the last inequality will be less than $\gamma/2$.

<u>lemma A.3</u>

Under assumptions 1)-5) in section 1 we have

a) $$n^{-1} \int_0^t [1 - \hat{S}(s)]^{-1} J_{13}(s) \; dN_{23}(s) \xrightarrow{P} (1-\lambda) \int_0^t g(s) \, ds$$

and

b) $$n^{-1} \int_0^t \hat{S}(s) [1 - \hat{S}(s)]^{-1} J_{13}(s) \; dN_{23}(s) \xrightarrow{P} (1-\lambda) \int_0^t S(s) \, g(s) \, ds$$

uniformly on [0,1].

<u>Proof of lemma A.3</u>

First notice that (by lemma A.2) part a) and b) of the lemma are equivalent, hence we shall only prove part a).

The only non-trivial condition in lemma A.2 is condition 6). Hence we must find appropriate A_n . To do this we will use lemma A.4 which is stated without proof.

For $t \le \varepsilon$

$$\hat{S}(t) = \prod_{s\le t} (1 - Y_1(s)^{-1} \, dN_{13}(s)) \le [1 - (\sup_{0\le s\le\varepsilon} Y_1(s))^{-1}]^{N_{13}(t)}$$

Hence by lemma A.4

$$[1 - \hat{S}(t)]^{-1} \le 2 \sup_{0\le s\le\varepsilon} Y_1(s) \, / \, N_{13}(t) \qquad \text{(A.1)}$$

if $\sup_{0\le s\le\varepsilon} Y_1(s) \ge 1$ and $\sup_{0\le s\le\varepsilon} Y_1(s) \ge N_{13}(t) - 1$.

The probability that these two conditions do not hold will tend to zero as n tends to infinity.

Thus for n larger than some N_2 we have with probabilty larger than $1-\delta/4$

$$\int_0^{\varepsilon} (1-\hat{S}(s))^{-1} \; J_{13}(s)\; n^{-1}\; dN_{23}(s) \leq 2n^{-1} \sup_{0\leq s\leq\varepsilon} Y_1(s) \int_0^{\varepsilon} N_{13}(s)^{-1} \; J_{13}(s)\; dN_{23}(s)$$

$$\leq 2 \int_0^{\varepsilon} N_{13}(s-)^{-1} \; J_{13}(s)\; dN_{23}(s) \;=\; A_n(\varepsilon)$$

which gives the definition of A_n. It now only remains to show that

$\lim_n EA_n(\varepsilon) \to 0$ as $\varepsilon \to 0$.

But $N_{13}(s-)^{-1} \; J_{13}(s)$ is left-continous and predictable in the family of σ-algebras $\{ \mathfrak{F}_t \; ; \; 0 \leq t \leq 1 \}$ and hence by lemma 2.1

$$A_n(t) \;=\; 2 \int_0^{t} N_{13}(s-)^{-1} J_{13}(s)\; (Y_0(s)+Y_2(s))\; F(s)\; \gamma(s)\; ds \;+\; M''(t)$$

where $M''(t)$ is a martingale in $\{ \mathfrak{F}_t \; ; \; 0 \leq t \leq 1 \}$

Consequently

$$E\, A_n(\varepsilon) \;=\; 2E \int_0^{\varepsilon} N_{13}(s-)^{-1} J_{13}(s)\; (Y_0(s)+Y_2(s))\; F(s)\; \gamma(s)\; ds \tag{A.2}$$

Further (by letting $p(s)=P(N_{i13}(s)=1)$ and omitting the argument s (and s-)) we have

$$E\, N_{13}^{-1}\, J_{13} = \sum_{k=1}^{n} k^{-1} \binom{n}{k} p^k (1-p)^{n-k} \leq 2 \sum_{k=1}^{n} (k+1)^{-1} \binom{n}{k} p^k (1-p)^{n-k}$$

$$\leq 2[p(n+1)]^{-1} \sum_{k=1}^{n} \binom{n+1}{k+1} p^{k+1} (1-p)^{n-k} \leq 2[p(n+1)]^{-1} \; . \tag{A.3}$$

And obviously

$$E\; (\; Y_0(s) + Y_2(s)\;) \;\leq n \; . \tag{A.4}$$

Hence by (A.2),(A.3),(A.4) we have

$$E\; A(\varepsilon) \quad \leq 4 \int_0^{\varepsilon} p(s)^{-1} \; F(s)\; \gamma(s)\; ds$$

We may write $p(t)=\int_0^{t} y_1(s)\, \alpha(s)\, ds$ and since $\inf_{0\leq t\leq 1} y_1(t) = y_1^0 > 0$ we find

$$p(t) \geq y_1^0 \int_0^{t} \alpha(s)\; ds = -\; y_1^0 \;\; \ln(1-F(t)) \geq y_1^0\; F(t)$$

Consequently we get

$$E\; A(\varepsilon) \leq (4/y_1^0) \int_0^{\varepsilon} \gamma(s)\; ds \;\to\; 0 \text{ as } \varepsilon \to 0.$$

Thus lemma A.3 is established.

Lemma A.4

If $y \geq m-1 \geq 0$, $y \geq 1$ and m is an positive integer we have

$$[\; 1 - (\; 1 - 1/y)^m \;]^{-1} \leq 2y/m \; .$$

The following lemma is useful in the derivation of the limit-process of $\sqrt{n}(\beta^*(t)-\beta(t))$ Many authors (e.g. Breslow&Crowley[1973]) have used and proved this result. They have however presented and proved the result in the course of an other proof, thus reducing the readability of these proofs.

Lemma A.5

Let $\{(U_n,V_n);\ n=1,2,\ldots..\}$ be a sequence of stochastic elements on the space D[0,1]xD[0,1] equipped with the Skorohod topology (See Billinglsey[1968]) that satifies

a) $V_n(0)=0$ for all n.

b) V_n is nondecreasing with probability 1.

c) $V_n(t) \overset{P}{\rightarrow} v(t)$ for some continous function v(t) uniformly on [0,1].

d) $U_n \overset{D}{\rightarrow} U$ for some U in D[0,1].

e) $P(\ U \in C[0,1]\ ) = 1$.

Then we have

$$X_n(t) = \int_0^t U_n(s)\ dV_n(s) \ \overset{D}{\rightarrow}\ X(t) = \int_0^t U(s)\ dv(s) \quad \text{on } D[0,1] \tag{A.5}$$

and

$$X_n(t) - \int_0^t U_n(s)\ dv(s) \overset{P}{\rightarrow} 0 \text{ uniformily on } D[0,1]\ . \tag{A.6}$$

Proof

The proof uses the Skorohod-Dudley-Theorem (See Gill[1980]). From this theorem we have that there exist a probability space with random elements $(U'_n\ ,\ V'_n\ )$ n=1,2,.... and $(U'\ ,\ v)$ such that $(U,v)\overset{D}{=}(U',v)$, a.s for all n $(U_n,V_n)=(U'_n,V'_n)$ and such that $(U'_n,V'_n) \overset{D}{\rightarrow} (U',v)$.

Now using lemma A.1 we get

$$X_n(t) = \int_0^t U'_n(s)\ dV'_n(s) \overset{a.s}{\rightarrow} \int_0^t U'(s)\ dv(s) \overset{D}{=} X(t)$$

which is simply (A.5). Further

$$X_n(t) - \int_0^t U_n(s)\ dv(s) \overset{D}{=} \int_0^t U'_n(s)\ dV'_n(s) - \int_0^t U'_n(s)\ dv(s)$$

$$\overset{a.s}{\rightarrow} \int_0^t U'(s)\ dv(s) - \int_0^t U'(s)\ dv(s) = 0$$

and this implies (A.6). Hence we have proved lemma A.5.

Proof of Theorem 4.1.

The proof of theorem 4.1 is in most respects a more worked-out version of the proof of proposition 4.1. The main differences can be stated in the following 3 points:

1) Lemma 4.1 may be invalid for $\varepsilon=0$ for certain choices of S(t) and g(t). Hence we will need an alternative version.
2) Allthough lemma 4.2 is true for $\varepsilon=0$ the proof of lemma 4.2 is not.
3) In the proof a functional Φ_ε is defined. Due to the alternative version of lemma 4.1 we must modify this functional and we have to be a bit more careful to establish the covergence.

First let

$$Z_1^{nq}(t) = \int_0^t (1-S(s))^{-q}\, dZ_1^n(s)$$

and

$$Z_2^{np}(t) = \int_0^t (1-S(s))^{p}\, dZ_2^n(s)$$

An useful way of stating the alternative to lemma 4.1 is then:

If $q < 0.5$ and $p > 0.5$ then

$$(Z_1^{nq} , Z_2^{np}) \xrightarrow{D} (Z_1^{q} , Z_2^{p}) \quad \text{on} \quad D[0,1]$$

where Z_1^q and Z_2^p are gaussian processes with expectation 0 and covariance functions

$$E\, Z_1^q(t)\, Z_2^q(u) = \int_0^{t\wedge u} (1-S(s))^{-2q}\, y_1(s)^{-1}\, \alpha(s)\, ds$$

and

$$E\, Z_2^p(t)\, Z_2^p(u) = (1-\lambda) \int_0^{t\wedge u} (1-S(s))^{2p-1}\, S(s)\, g(s)\, ds$$

We omit the proof of this result, but the MCLT as stated in Andersen&Gill[1982] is appliciable.

Next we turn to the extension of lemma 4.2 to the case $\varepsilon=0$. The amendments that must be done lies the the proofs of (4.9) and (4.10).

To see that (4.9) hold when $\varepsilon=0$ as well let

$$X_n^p(t) = (1-S(t))^p\, [\, Z_2^n(1) - Z_2^n(t)\,]$$

Then we may write

$$V_{n2}^*(t)-V_{n2}(t) = \int_0^t X_n^p(s)(1-S(s))^{-p}\, [(n/Y^*(s))Y_1(s)^{-1}dN_{13}(s)-y^*(s)^{-1}\alpha(s)ds]$$

But by lemma A.2 we have for $p < 1$

$$\int_0^t (1-S(s))^{-p} \left[(n/Y^*(s))Y_1(s)^{-1} dN_{13}(s) - y^*(s)^{-1}\alpha(s)ds \right] \xrightarrow{P} 0 \text{ on } D[0,1]$$

By lemma A.5 it will then suffice to show that there exist p (< 1) such that $X_n^p(t)$ will converge in distribution to a process $X^p(t)$ with $P(X^p(t) \in C[0,1]) = 1$. To show that a such p does exist it is essential that we have established the limit distribution of $Z_2^{np}(t)$ (for p>0.5). For these values of p $X^p(t)$ will be a gaussian process with continous covariance function, hence on C[0,1] with probability 1. To see this note

$$X_n^p(t) = (1-S(t))^p \int_t^1 (1-S(s))^{-p} \, dZ_2^{np}(s)$$

By theorem 2.4.3 in Gill[1980] it is now sufficient to show that if $z_n(t)$ is a sequence of functions on D[0,1] and z(t) a continous function such that $\sup_{0\le t\le 1} | z_n(t) - z(t) | \to 0$ then also

$$\sup_{0\le t\le 1} (1-S(t))^p \left| \int_t^1 (1-S(s))^{-p} \, d\,(z_n(s)-z(s)) \right| \to 0 . \qquad \text{(A.7)}$$

But by partial integration

$$\begin{aligned}\int_t^1 (1-S(s))^{-p} \, d(z_n(s) - z(s)) \; = \; & (z_n(1) - z(1)) \, (1-S(1))^{-p} \\ - \; & (z_n(t) - z(t)) \, (1-S(t))^{-p} \\ + \; & \int_t^1 (z_n(s) - z(s)) \, d(1-S(s))^{-p}\end{aligned}$$

and hence the expression in (A.7) is dominated by

$$2 \sup | z_n(t) - z(t) | \to 0 \text{ as } n \to \infty .$$

Consequently (A.7) hold, yielding that (4.9) hold for $\varepsilon=0$ as well.

To show that also (4.10) is valid for $\varepsilon=0$ we will show that each of the four terms on the right-handside in (4.13) will tend to zero (in probability on D[0,1]). For all terms arguments simiiliar to the one above will be applied.

First term in (4.13): Due to lemma A.1 and to

$$\int_0^t (1-S(s))^{-p} \, (n/Y^*(s)) \, Y_1(s)^{-1} dN_{13}(s) \xrightarrow{P} \int_0^t (1-S(s))^{-p} y^*(s)^{-1} \alpha(s)ds$$

on D[0,1] for p<1 it will be sufficent to find a p such that

$$(1-S(s))^p \int_s^1 \frac{S(r)}{1-S(r)} \, Z_1^n(r) \left[(1-\lambda)g(r)dr - \frac{J_{12}(r)dN_{23}(r)}{n(1-S(r))} \right] \xrightarrow{P} 0$$

D[0,1].

But by partial integration we see that this will follow from
$$\int_s^1 S(r)\ (1-S(r))^{p-1}\ Z_1^n(r)\ [\ (1-\lambda)g(r)dr - \frac{J_{13}(r)dN_{23}(r)}{n(1-S(r))}\ ] \xrightarrow{P} 0$$
on D[0,1]. But by partial integration, letting $q=1-p$, we get that $(1-S(r))^{-q}\ Z_1^n(r)$ will converge to a process that is element in C[0,1] with probability 1 if the corresponding process $Z_1^{nq}(r)$ has this same property. Hence choosing $p\varepsilon\langle 0.5,1\rangle$ we may apply lemma A.5 to get the wanted conclusion.

Second term in (4.13): By a argument similiar to the preceeding we have that for $p>0.5$ will
$$(1-S(s))^p \int_s^1 \frac{S(r)}{1-S(r)}\ Z_1^n(r)\ \frac{J_{13}(r)\ dN_{23}(r)}{n\ (1-S(r))}$$
converge in distribution to a process that is on C[0,1] with probability 1. Since also (for $p<1$)
$$\int_0^t (1-S(s))^{-p}\ [\ y^*(s)^{-1}\ \alpha(s)ds - (n/Y^*(s))\ Y_1(s)^{-1}dN_{13}(s)\ ] \xrightarrow{P} 0$$
on D[0,1] the second term in (4.13) will "disappear" also for $\varepsilon=0$.

Third term: It is possible to show that for $p\varepsilon\langle 0.5,1\rangle$ will
$$Z_1^n(r)\ (1-S(r))^p\ [\ \frac{J_{13}(r)}{1-S(r)} - \frac{J_{13}(r)}{1-\hat{S}(r)}\ ] \xrightarrow{P} 0 \quad \text{on} \quad D[0,1] \qquad (A.8)$$
This is done using the the tecnique of partial integration to show that this will follow from
$$\int_0^s (1-S(r))^p\ [\ \frac{J_{13}(r)}{1-S(r)} - \frac{J_{13}(r)}{1-\hat{S}(r)}\ ]\ dZ_1^n(r) \xrightarrow{P} 0 \quad \text{on } D[0,1]$$
and then the MCLT to show that the above expression is true. But by partial intgration again we get that (A.8) is sufficient to show that the third term tends to zero in probability on D[0,1].

Fourth term in (4.13): As with the third term it will be sufficient to show that for $p\varepsilon\langle 0.5,1\rangle$ will (on D[0,1])
$$[\ -S(r)\ Z_1^n(r) - \sqrt{n}\ (\hat{S}(r)-S(r))\ ]\ (1-S(r))^p\ \frac{J_{13}(r)}{1-\hat{S}(r)} \xrightarrow{P} 0$$
This can be done (essentially) the same way as with (A.8).

It now remains only to show that the process $V_n(u)$ will converge to the process $V(u)$ defined in theorem 4.1. Note that we may write

$$\begin{aligned} V_n(u) &= \int_0^u (1-S(t))^q \, dZ_1^{nq}(t) \\ &+ \int_0^u \int_t^1 (1-S(s))^{-p} \, dZ_2^{np}(s) \; y^*(t)^{-1} \, \alpha(t)\, dt \\ &- (1-\lambda) \int_0^u \int_t^1 \frac{S(s)}{1-S(s)} \int_0^s (1-S(r))^q \, dZ_1^{nq}(r) \, g(s)ds \; y^*(s)^{-1} \alpha(s)ds \end{aligned}$$

We will define a functional $\Psi : D[0,1]^2 \rightarrow D[0,1]$ by

$$\Psi(Z_1^{nq} , Z_2^{np}) (u) = V_n(u)$$

We then have for $q \varepsilon \langle 0,0.5\rangle$ and $p \varepsilon \langle 0.5,1\rangle$ that on $D[0,1]$

$$\Psi(Z_1^{nq} , Z_2^{np}) (u) \xrightarrow{D} \Psi(Z_1^{q} , Z_2^{p}) (u) \tag{A.9}$$

This is true because

1) $(Z_1^{nq} , Z_2^{np}) \xrightarrow{D} (Z_1^{q} , Z_2^{p})$ on $D[0,1]^2$ for these values of (q,p) and with the limit-element in $C[0,1]^2$ with probability 1.
2) if (z_1^n , z_2^n) is a sequence of fucunctions converging uniformly to (z_1 , z_2) on $[0,1]^2$ then also $\Psi(z_1^n, z_2^n)$ will converge uniformly to $\Psi(z_1^n, z_2^n)$ on $[0,1]$.

That 1) and 2) imply (A.9) follows from theorem 2.4.3 in Gill[1980]. That 2) indeed hold may be shown by performing integrations.

$\Psi(Z_1^q , Z_2^p) (u)$ will obviously be gaussian and that it has the same covariance-function as $V(u)$ may be demonstrated the same way way as we derived the covariance-function of $\Phi_\varepsilon(Z_1, Z_{2\varepsilon})$ (in the proof of proposition 4.1). Hence the proof of theorem 4.1 is comcomplete and the paper is finished (PUH).

References.

Aalen O.O. [1976] : "Nonparametric inference in connection with multiple decrement models." Scand.J.Stat. 3, 15-27

Aalen, O.O. [1978a] : "Nonparametric estimation of partial transition probabilities in multiple decrement models." Ann.Stat. 6, 534-545

Aalen, O.O. [1978b] : "Nonparametric inference for a family of counting processes." Ann.Stat. 6, 701-726

Aalen, O.O. & Johansen, S. [1978] : "An empirical transition matrix for non-homogeneous Markov-chains based on censored observations." Scand.J.Stat. 5, 141-150

Andersen & Gill [1982] : "Cox's regression model for counting processes: A large sample study." Ann.Stat. 10, 1100-1120

Billingsley, P [1968] : "Convergence of probability measures." New York: Wiley.

Breslow & Crowley [1974] : "A large sample study of the life table and product-limit estimates under random censorship." Ann.Stat. 2, 437-453

Dempster, A.P., Laird, N.M., & Rubin, D.B. [1977] : "Maximum likelihood from incomplete data via the E-M-algorithm (with discussion.)" J.Roy.Stat.Soc. 39,1-37

Efron, B. [1967] : "The two sample problem with censored data." Proc.Fifth Berkeley Symp. Math. Stat. Prob. 4, 831-853

Gill, R.D. [1980] : "Censoring and stochastical integrals." Math. Centre Tracts 124, Mathematical Centre,Amsterdam.

Kaplan, E.L. & Meier, P. [1958] : "Nonparametric estimation from incomplete observations." J.A.S.A. 53, 457-481

Peto, R. [1973] : "Experimental survival curves for interval-censored data." Applied Statistics 22, 86-91

Samuelsen, S.O. [1984] : "Ikke-Parametrisk Estimering av Fordelingsfunksjoner når Datamaterialet er Venstre- eller Dobbelsensurert. Asymptotisk Teori." Unpublished Graduate Thesis (in Norwegian). University of Oslo.

Turnbull, B.W. [1974] : "Nonparametric estimation of a survivorship function with dubly censored data." J.A.S.A. 69,169-173

Turnbull, B.W. [1976] : "The empirical distribution function with arbitrary grouped, censored and truncated data." J.Roy.Stat.Soc.B 38, 290-295